

Methoden zur genomischen Zuchtwertschätzung: Aktueller Stand und Zukunft

C. Stricker

stricker@genetics-network.ch
agn Genetics GmbH, Davos



2 Ansätze

- ▶ Klassischer, zweistufiger Ansatz
 - ▶ QTL detection (linkage, GWAS), multiples Testen in kleinen Datensets
 - ▶ false positives: Loci ohne Effekte 'signifikant'
 - ▶ false negatives: QTL nicht signifikant
 - ▶ MAS, false negatives → begrenzen Genauigkeit.
 - ▶ missing heritability problem
- ▶ Genomweite Evaluation
 - ▶ einstufiger Ansatz, kein Hypothesen testen
 - ▶ alle Marker im Model, Variablenselektion
 - ▶ $n \ll \ll p$: a priori Annahmen über Parameter des Modells (Bayes Statistik)



A unified Bayesian Approach

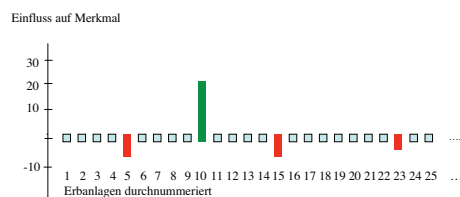
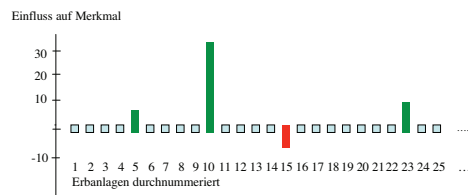
$$y_i = \mu + \sum_j x_{ij} a_j + e_i$$

- ▶ Locus j im Modell mit Wahrscheinlichkeit $1 - \pi$, π ist Parameter des Modells
- ▶ $a_j \sim N(0, \sigma_j^2)$ oder $\sim t(\nu)$ with $\nu > 4$ als Parameter im Modell
- ▶ $\sigma_j^2 \sim$ scaled-inverse $\chi^2(\nu)$, $\nu > 4$, Parameter des Modells
- ▶ $\sigma_e^2 \sim$ scaled-inverse $\chi^2(\nu)$, $\nu > 4$, Parameter des Modells
- ▶



Praktische Interpretation

- ▶ Zuchtwert $\sum_j x_{ij} a_j$



Praktische Interpretation

sample	#SNPs/cM				
	1	5	10	15	20
1	0.96/0.21	0.98/0.34	0.98/0.43	0.98/0.47	0.98/0.50
2	0.96/0.20	0.98/0.32	0.98/0.41	0.98/0.46	0.98/0.47
3	0.96/0.19	0.98/0.33	0.98/0.42	0.98/0.45	0.98/0.46
4	0.97/0.24	0.98/0.42	0.98/0.51	0.98/0.56	0.98/0.59

Table: Bestimmtheitsmasse Trainings-/Validierungsdaten aus Simulationsstudie ASR: Merkmal mit $h^2 > 0.95$, 30 Chromosomen, effektive Populationsgrösse 100, 2120 Beobachtungen in training data, 2000 Nachkommen der Trainingsdaten als Validierungsdaten.



Praktische Interpretation

- ▶ Woher kommt die Genauigkeit genomischer Zuchtwerte: Cosegregation oder Linkage Disequilibrium?
 - ▶ falls Cosegregation → populationspezifisch, häufiges re-Training notwendig, variable selection models not necessary
 - ▶ falls LD → populationsübergreifend, crossbreeding, kein re-Training, nur bei Neumutationen und Fixierung von QTL, variable selection models, $n \ll p$ problem, limited bayesian learning)
- ▶ wie erreichen wir höheres LD → high density SNP panels

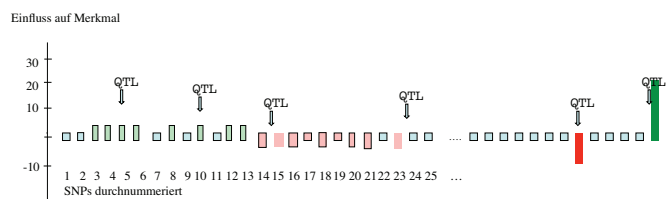
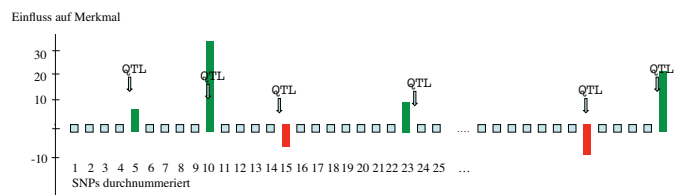


Zukünftige Entwicklungen

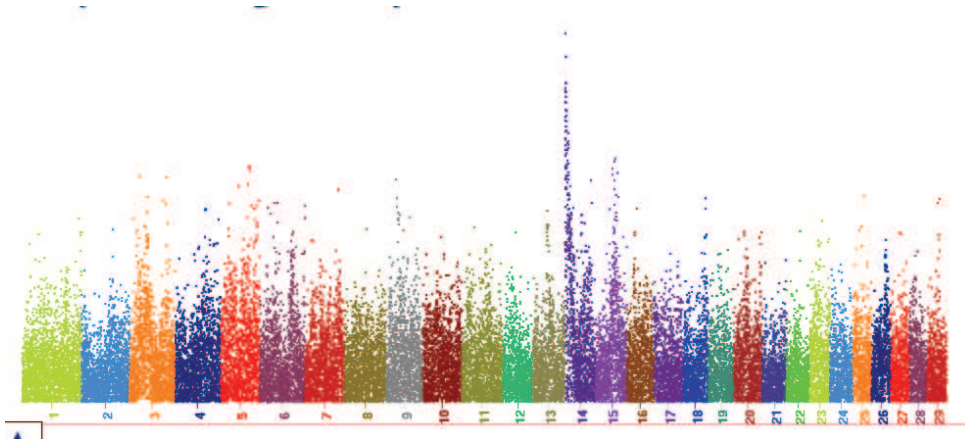
- ▶ in der Praxis
 - ▶ NZP ↓
 - ▶ Wartestiere, Altstiere ↓
 - ▶ Genetischer Fortschritt pro Jahr ↑
 - ▶ neue Phänotypen
- ▶ in der Technik/Theorie
 - ▶ 800k Chip, Sequenzdaten
 - ▶ QTL Mapping: Chromosomale Segmente anschauen.



Zukünftige Entwicklungen



Zukünftige Entwicklungen



Zukünftige Entwicklungen

Marker	Effect	EffectVar	ModelFreq	GeneFreq
30576	-2.27E+02	5.28E+04	0.9597	0.312
48684	2.52E+01	3.01E+03	0.2742	0.197
39870	1.80E+01	1.63E+03	0.2501	0.411
8948	1.58E+01	1.14E+03	0.2456	0.596
21041	1.78E+01	1.87E+03	0.2059	0.859
48094	1.58E+01	1.63E+03	0.1978	0.282
47936	-1.14E+01	8.99E+02	0.1859	0.614
34950	-1.16E+01	9.40E+02	0.1833	0.501
51741	6.50E+00	3.98E+02	0.1693	0.683
53381	8.11E+00	5.15E+02	0.1577	0.47
33245	8.23E+00	5.42E+02	0.1483	0.327
18161	-7.53E+00	6.01E+02	0.1442	0.366

Zukünftige Entwicklungen

Marker	abs(Effect)	Effect	EffectVar	ModelFreq	GeneFreq
30576	227	-2.27E+02	5.28E+04	0.9597	0.312
48684	25.16	2.52E+01	3.01E+03	0.2742	0.197
39870	17.96	1.80E+01	1.63E+03	0.2501	0.411
21041	17.78	1.78E+01	1.87E+03	0.2059	0.859
48094	15.82	1.58E+01	1.63E+03	0.1978	0.282
8948	15.76	1.58E+01	1.14E+03	0.2456	0.596
33302	11.89	1.19E+01	2.14E+03	0.1011	0.058
34950	11.62	-1.16E+01	9.40E+02	0.1833	0.501
47936	11.37	-1.14E+01	8.99E+02	0.1859	0.614
30571	8.748	-8.75E+00	1.71E+03	0.0708	0.342
33245	8.233	8.23E+00	5.42E+02	0.1483	0.327
15212	8.132	-8.13E+00	7.30E+02	0.1408	0.686



Zukünftige Entwicklungen

- ▶ an 100 Stelle : Effekt 2.78 Modelfreq 0.076
- ▶ an 1000 Stelle : Effekt 0.92 Modelfreq 0.042
- ▶ an 2000 Stelle : Effekt 0.6581 Modelfreq 0.036



